DTIC FILE COPY

(2)

AD-A215 462

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

DTIC ELECTE OCT 26 1989 B

1a REPORT SECURITY CLASSIFICATION	1b RESTRICTIVE MARKINGS NA
2a SECURITY CLASSIFICATION AUTHORITY NA	3 DISTRIBUTION/AVAILABILITY OF REPORT Distribution Unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) Wayne State University Detroit, Michigan 48201	5 MONITORING ORGANIZATION REPORT NUMBER(S) NA
6a NAME OF PERFORMING ORGANIZATION Wayne State University; 6b OFFICE SYMBOL (If applicable) NA	7a NAME OF MONITORING ORGANIZATION Office of Naval Research
6c ADDRESS (City, State, and ZIP Code) Department of Biochemistry School of Medicine 540 E. Canfield St. Detroit, Michigan 48201	7b ADDRESS (City, State, and ZIP Code) 800 N. Quincy St. Arlington, Virginia 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research; 8b OFFICE SYMBOL (If applicable) ONR	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-87-K-0081
8c ADDRESS (City, State, and ZIP Code) 800 N. Quincy St. Arlington, Virginia 22217-5000	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO 61153N; PROJECT NO RR04106; TASK NO 441-3031; WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
Evolution and Analysis of the Functional Domains of the Chimeric Proteins that Initiate Pyrimidine Biosynthesis

12 PERSONAL AUTHOR(S)
Evans, David Robert

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Annual	FROM 8/1/88 TO 8/31/89	October 15, 1989	6

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB GROUP	Chimeric or multifunctional proteins, protein domains, pyrimidine biosynthesis; carbamyl phosphate synthetase, aspartate transcarbamylase, dihydroorotase, molecular evolution.
06	03		

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The objective of this project is to test the hypothesis that the complex proteins with novel functions arose in the course of evolution by combining structural domains having partial functions. The research focuses on the enzymes that catalyze *de novo* pyrimidine biosynthesis. While the reactions are the same in most organisms, there are striking differences in the structure and regulation of these enzymes. This year we completed the sequence of the mammalian protein CAD, a 243 kDa polypeptide which carries glutamine dependent (GLN) carbamyl phosphate synthetase (CPS), aspartate transcarbamylase (ATC) and dihydroorotase (DHO) activities. Phylogenetic analysis suggests that the mammalian chimeric protein was formed by stepwise translocation and fusion of ancestral genes that occurred prior to the major radiation that lead to fungi, plants and animals. The sequence divergence suggests that the fused and monofunctional DHO domains have a different evolutionary history. In contrast, sequence comparisons and molecular modeling shows that the ATC domain is a trimer of 34 kDa domains that has been highly conserved thoughout the course of evolution. This conclusion conflicts with the structural studies of the prokaryotic class C ATCases which were thought to have a radically different structural organization.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION (U)
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Michael Marron	22b TELEPHONE (Include Area Code) (202) 696-4760; 22c OFFICE SYMBOL ONR

DD Form 1473, JUN 86 Previous editions are obsolete SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

89 120

19. ABSTRACT (con't)

This apparent contradiction was resolved by reexamining the subunit structure of *P. fluorescens* ATCase. While much larger that other prokaryotic ATCases, this class C enzyme consists of six copies of a 34 kDa catalytic polypeptide and six copies of a 45 kDa polypeptide that probably mediates regulation. Since the major translocations and fusions appear to predate the emergence of eukaryotic organisms, we have begun to examine the structural organization of the pyrimidine specific genes in archebacteria. A separate pyrimidine specific CPSase was cloned from *Methanosarcina barkeri*. The gene duplication leading to separate CPSases for pyrimidine and arginine pathways has occurred in this archebacteria but the genes encoding GLN and CPS domains are not fused.

ANNUAL REPORT ON CONTRACT N00014-87-K-0081 R&T CODE 4413031

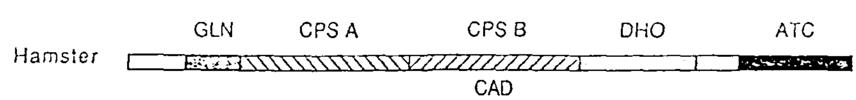
PRINCIPAL INVESTIGATOR: David R. Evans

CONTRACTOR: Wayne State University, Detroit, Michigan

CONTRACT TITLE: The Evolution and Analysis of the Functional Domains of the Chimeric Proteins that Initiate Pyrimidine Biosynthesis.

PROJECT PERIOD: August 1, 1988 - August 31, 1989

RESEARCH OBJECTIVE: To determine the structural organization and trace the evolutionacry development of the complex multi-domain proteins involved in *de novo* pyrimidine biosynthesis. The enzymes which catalyze the first three steps in the pathway, glutamine dependent (GLN) carbamyl phosphate synthetase (CPS), aspartate transcarbamylase (ATC) and dihydroorotase (DHO) are separate proteins in eubacteria, but are consolidated in a single 243 kDa chimeric plypeptide in mammals and other higher eukaryotes. We have shown previously that the polypeptide is organized into at least five functional domains.



PROGRESS REPORT.

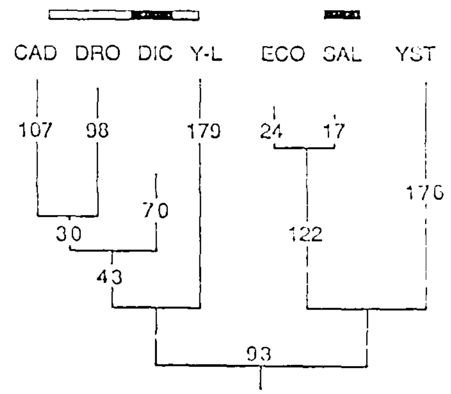
Origin of the Trifunctional Polypeptide

During the last year we used primer extension to construct a cDNA molecule (Bein, Simmer and Evans, manuscript in preparation) corresponding to the 5' end of the CAD mRNA. This clone allowed us to complete the sequence determination of the entire CAD protein. We are now analyzing the phylogenetic relationship of each of the domains using sequence data of several prokaryotic and eukarotic enzymes including the hamster protein (CAD) and partial sequence data for the pyrimidine biosynthetic complex of Dictyostelium discoideum (DIC), published this year (Faure et al., Eur. J. Biochem. 179: 345, 1989).

While the GLN, CPS and ATC domains are clearly homologous (approximately 50% sequence identify) to the corresponding monofunctional *E. coli* enzymes, the mammalian

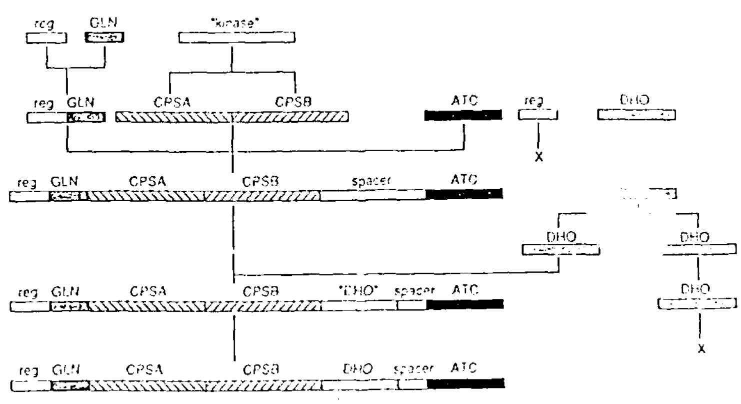
DHO domain is strikingly different than its bacterial counterpart (Simmer et al., PNAS, in press) . Another surprising discovery, was that the mammalian DHO domain is homologous to the the long interdomain linker region (Y-L) that connects the CPS and ATC domains in yeast but very different than the functional yeast DHOase (YST) which is encoded by a separate gene. The phylogenetic analysis shows that the mono-functional and fused DHOases have a different evolutionary history since the dendrogram does not conform to the accepted phylogeny of the organisms represented. Although the *Dictyostelium* lineage predates the major radiation that lead to fungi, plants and animals, the dendrogram clusters the enzyme from *Dictyostelium* with higher eukaryotes and the yeast enzymes with the prokaryotic DHOases.

RELATIONSHIP OF DHO SEQUENCES



The disparate sequences of the two class of DHOases can be plausibly explained by divergent evolution following duplication of an ancestral gene (scheme shown below). According to this model, the fusion of CPSase and ATCase genes, separated by an approximately 300 bp spacer, occurred sometime between the divergence of bacteria and slime molds. This early event was followed by duplication of a monofunctional DHOase gene, one copy of which was translocated and inserted into the spacer region. Perhaps initially non-functional ("DHO"), reactivation (DHO) in the *Dictyostelium* and metazoan lineages, with the concomitant advantages of coordinate regulation, lead to the extinction of the monofunctional dihydroorotase. Reactivation did not occur in yeast and the separate monofunctional DHOase was preserved.

ORIGIN OF THE PYRIMIDINE BIOSYNTHETIC COMPLEXES



If this explanation (developed more fully in Simmer et al., PNAS in press) is correct all of the dihydroorotases are descendants of a common ancestor and the sequence differences between the two families is a consequence of differences in structural constraints imposed on the fused and monofunctional dihydroorotases.

In contrast, the sequence and structure of the fused and monofunctional aspartate transcarbylase domains are highly homologous. We built an energy minimized model of the mammalian ATCase domain using the x-ray coordinates of the *E. coli* enzyme as a tertiary template (Scully and Evans, manuscript in preparation). Favorable hydrophobic interactions, a compact globular shape and a normal distribution of hydrophobic and hydrophillic side chains suggests that the model is a plausible representation of the structure of the mammalian ATCase domain. The backbone carbonyl carbons are nearly superimposable, the active site regions are virtually identical and the trimeric subunit contacts are similar.

The following conclusions are consistent with these studies,

1. Evolution of the multidomain occurred by stepwise translocation and fusion of ancestral genes coding for monofunctional proteins.

2. The formation of a large chimeric protein was an early event which predates the radiation which lead to the major families of eukaryotic organisms.

3. The sequence of most domains has been highly conserved (eg. ATCase) while the DHO domain has undergone much more extensive divergence perhaps indicative of adaptive changes.

The major gene fusions and translocations appeared to have occurred much earlier in the course of evolution than was generally believed when we begun this study. Thus we are now focusing our attention on more primitive organisms, the bacteria and archebacteria.

Prokaryotic Class A Aspartate Transcarbamylases

Three classes of ATCase, which differ in strucutral organization and regulatory properties, have been identified in eubacteria. *E. coli* aspartate transcarbamylase, a well characterized class B enzyme, is a dodecamer of two catalytic subunits and three regulatory subunits. The catalytic subunit, a trimer of 34 kDa catalytic chains, is catalytically active but is not regulated, while the regulatory subunit, a dimer of 17 kDa regulatory chains, binds allosteric effectors but is inactive. *B. subtilis* ATCase, a typical class C enzyme, is an un-unregulated trimer of identical 34 kDa catalytic chains. The structure and regulation of the class A enzymes, the largest aspartate transcarbamylases, are not well understood.

Our studies have focused on *P. fluorescens* ATCase, a class A ATCase reported to be a dimer of two identical 180 kDa subunits. The enzyme is inhibited by broad range of nucleotides but does not exhibit cooperative substrate binding and thus apparently has a less sophisticated mode of regulation than the *E. coli* enzyme.

The large size of the molecule raised the possibility that the *P. fluorescens* enzyme was a multidomain protein analogous to the mammalian complex. To determine whether either of CPS or DHO activities are associated with *P. fluorescens* ATCase, cell extracts were fractionated on Sephacyrl S-300 chromatographic column calibrated with a proteins of known molecular weight. This experiment showed that all of enzymes are separate proteins. The molecular weights of CPS and DHO are 160 kDa and 86 kDa respectively, comparable to the size of the *E. coli* enzymes, while the ATCase is much larger.

The material we have isolated represents the purest active enzyme preparation thus far obtained. Contrary to previous reports we found that the molecule consists of two subunits, 34 kDa and 45 kDa, which are present in approximately stoichiometric amounts in the complex. We believe that the catalytic activity is associated with the 34 kDa polypeptide.

Preliminary protein sequence data suggest that this protein is homologous with the *E. coli* catalytic subunit.

We were concerned that these polypeptides were degradation products of a large multidomain polypeptide so we developed a new autoradiographic assay for assessing proteolytic activity in cell extracts. The assay uses radiolabelled CAD as a substrate (one of the inter domain linker regions in CAD is extraordinarily senstive to a broad range of proteases). We conclude that proteolysis is not occurring under the conditions that the isolation was carried out.

Sedimentation velocity and gel filtration studies showed that the enzyme is a stable complex with a molecular weight of approximately 480,000 and a Stokes radius of 77 Å. Thus the enzyme is a dodecamer composed of six copies of each of the two types of subunit. We are now attempting to dissociate the complex and isolate the constituent subunits by gel filtration, so that the structure and function of each polypeptide can be fully characterized.

In summary, while confirmatory experiments remain to be carried out, we can tentatively draw the following conclusions:

1) The first three steps of pyrimidine biosynthesis in *P. flourescens* are carried by separate proteins, and while the CPSase and DHOase are similar to the enzymes found in *E. coli*, the ATCase is unusual. It is not however a dimer of 180 kDa polypeptides, but rather a duodecamer comprised of two types of polypeptide chain.

2) The catalytic subunit is probably a trimer of identical 34 kDa polypeptides such as that found in other prokaryotic and eukaryotic ATCases supporting the hypothesis that they have all evolved from a highly conserved ancestral domain.

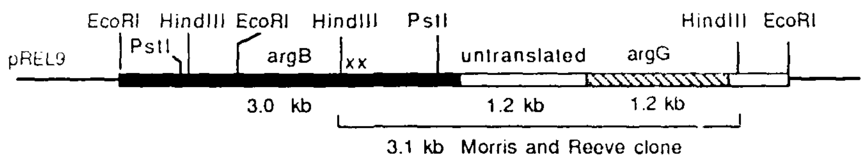
3) The function of the second polypeptide is unknown. While it is likely to be responsible for regulation, its large size, nearly three times the mass of the *E. coli* enzyme regulatory chain, suggest that it may have other functions. Identification of these functions may provide a clue to the origin of the highly specialized regulatory subunit found in *E. coli*.

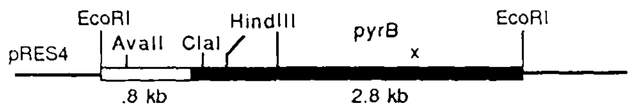
These studies (Bergh and Evans, manuscript in preparation) confirm that the class A aspartate transcarbamylases are very large proteins, but completely revise previous ideas regarding the structural organization of these molecules.

Archebacteria

We have begun to look at the structural organization of the genes encoding the pyrimidine biosynthetic enzymes. These studies have been aided by investigators that kindly shared libraries and other materials: lambda libraries of *Methanoacrcina acetivorans* (Dr. Kevin Sowers and Robert Gunsalus, UCLA); *Thermoplasma acidophilum* (Dr. Charles Daniels, Ohio State); cosmid blots and selected clones of *Halobacterium volcanii* (Drs. W. Ford Doolittle, Robert Charlebois and Leo Schalkwyk).

In addition Drs. John Reeve and Joseph Krzycki (Ohio State University) have generously sent us *Methanosarcina barkeri* MS vectors and cells. We divised a method for isolating the *M. barkeri* DNA after our initial attempts resulted in massive shearing and used the DNA to construct a lambda genomic library. Morris and Reeve (J. Bacteriol. 170: 3125, 1988) had cloned the argininosuccinate synthetase gene (argG) and found that one of their clones contained a 1.2 kb sequence that was homologous to carB, the gene that encodes the large subunit of *E. coli* CPSase. We screened our library using two end labelled synthetic oligonucleotides complimentary to highly conserved regions of the CPSase as probes. Plaque hybridization revealed two classes of clones. The strongest signals were obtained from phage containing fragments of the CPSase gene previously discovered by Morris and Reeves. These plaques hybridize to both probes (xx, see diagram below). We have now isolated clones (eg. pREL9 below) that encode all or very nearly all of this CPSase gene (argB). Plaques giving weaker, but still definitely positive, signals were also subcloned (eg. pRES4 below) analyzed. Restriction mapping and Southern analysis showed that these were derived from a distinctly different gene which hybridizes to only one of the two probes (x). Sequencing studies now underway are confirming that this gene codes for an entirely separate CPSase which we have desginated pyrB.





Given its close proximity to the argG gene is is likely that the CPSase discovered by Morris and Reeves is specific for arginine biosynthesis, while the CPSase gene (pyrB) we have found is the pyrimdine specific enzyme. Unlike the situation in *E. coli*, the gene duplication which gave rise to two separate carbamyl phosphate synthetases specialized for arginine and pyrimidine biosynthesis has occurred in *M. barkeri*. Further analysis of these clones should give the first information about the structural organization of the pyrimidine pathway enzymes in archebacteria. Positive clones have also been tentatively identified in the *H. volcanii* and *M. acetivorans* libraries, but these have not yet been characterized.

Our efforts are now being directed towards more fully characterizing the putative pyrimidine specific CPSase and in mapping the location of the other pyrimidine specific enzymes in Methanosarcina barkeri and the other archebacteria.